AI for radiographic COVID-19 detection selects shortcuts over signal

Year: 2020 | Citations: 594 | Authors: A. DeGrave, Joseph D. Janizek, Su-In Lee

Abstract

Artificial intelligence (AI) researchers and radiologists have recently reported AI systems that accurately detect COVID-19 in chest radiographs. However, the robustness of these systems remains unclear. Using state-of-the-art techniques in explainable AI, we demonstrate that recent deep learning systems to detect COVID-19 from chest radiographs rely on confounding factors rather than medical pathology, creating an alarming situation in which the systems appear accurate, but fail when tested in new hospitals. We observe that the approach to obtain training data for these AI systems introduces a nearly ideal scenario for AI to learn these spurious ‘shortcuts’. Because this approach to data collection has also been used to obtain training data for the detection of COVID-19 in computed tomography scans and for medical imaging tasks related to other diseases, our study reveals a far-reaching problem in medical-imaging AI. In addition, we show that evaluation of a model on external data is insufficient to ensure AI systems rely on medically relevant pathology, because the undesired ‘shortcuts’ learned by AI systems may not impair performance in new hospitals. These findings demonstrate that explainable AI should be seen as a prerequisite to clinical deployment of machine-learning healthcare models. The urgency of the developing COVID-19 epidemic has led to a large number of novel diagnostic approaches, many of which use machine learning. DeGrave and colleagues use explainable AI techniques to analyse a selection of these approaches and find that the methods frequently learn to identify features unrelated to the actual disease.